基于相似性群体的混合型 Web 服务推荐*

谢　琪[1,2]　崔梦天[1]
[1](西南民族大学计算机科学与技术学院　成都 610225)
[2](香港中文大学计算机科学与工程学系　香港 999077)

摘要:【目的】改进服务计算环境下 Web 服务推荐数据稀疏性导致的相似服务或相似用户缺失的问题。【方法】根据相似性距离分别为目标用户和服务构造个性化的相似性用户群体和服务群体, 同时使用用户和服务群体中心之间的群体相似性设计新的混合型推荐算法(GHQR)。【结果】使用 197 万条真实 Web 服务质量数据集的实验结果表明, 与 UPCC 和 IPCC 两种推荐算法相比, GHQR 的标准平均绝对误差(NMAE)平均下降 31%、69%, 覆盖率平均提高 105%、163%。【局限】实验仅对服务质量属性响应时间进行分析, 还需对其他 Web 服务质量属性如吞吐率等进行验证。【结论】与 WSRec 和 CFBUGI 推荐算法相比, GHQR 的 NMAE 平均下降 26%、7.7%, 覆盖率平均提高 188%、4%。GHQR 不仅能提高预测的准确性, 而且覆盖率也获得显著提高。

关键词: 个性化推荐　服务计算　Web 服务　服务质量
分类号: TP311

1　引　言

随着互联网上 Web 服务的不断增加, 如何为服务用户提供满足其需求的Web服务已成为服务计算领域内的一个研究热点和重点[1-9]。为了适应 Web 服务的动态网络环境, 基于服务质量(Quality of Service, QoS)的 Web 服务预测方法已成为该问题的主要解决方案。由于 Web 服务的 QoS 值会受到网络延迟、用户/服务物理位置等因素的影响, 可能使得 Web 服务的 QoS 值并不是服务提供商所提供的 QoS 值[2]。不同的服务用户调用同一个 Web 服务可能具有不同的 QoS 值。如日本 dyndns.org 服务提供商提供的语音验证码服务 CaptchaAudio, 对于在美国的用户 u_1 调用该服务获得的响应时间为 0.222 秒, 而对于在日本的用户 u_3 调用该服务获得的响应时间为 0.358 秒。因此如何获得动态实时的真实 QoS 值是 Web 服务推荐亟待解决的关键问题。

目前基于 QoS 的推荐研究主要集中于将传统协同过滤的推荐算法[10-13]应用于 Web 服务推荐[1-2, 5-6]。邵凌霜等[1]较早地提出一种基于 QoS 的 Web 服务推荐算法, 该算法将传统协同过滤中的商品评分值替换为服务用户调用服务的 QoS 值。Zheng 等[2]提出一种基于用户和基于服务的混合型 Web 服务推荐算法, 并发布了真实 Web 服务质量数据集 WS-DREAM。张璇等[5]考虑可信因素对服务推荐的影响, 提出一种基于协同过滤的可信 Web 服务推荐方法。王海艳等[6]引入服务的推荐属性特征, 并对传统的相似度计算公式进行改进, 从而提出一种基于可信联盟的服务推荐方法。姜波等[7]将 Web 服务之间的关系抽象成一个二部图, 提出一种基于二部图的服务推荐算法。虽然上述研究针对单个用户或单个服务对传统推荐算法进行改进, 并获得优于传统推荐算法的推荐性能, 但这些研究较少考虑到用户或服务所在的用户群体或服务群体对 Web 服务推荐结果的影响, 其推荐性能仍需获得进一步的提高。以用户群体或服务群体为单位进行推荐不仅能缓解因数据稀疏性导致的相似邻居缺失问题, 而且能更准确地反映出用户之间或服务之间的真实相似性, 从而提

通讯作者: 谢琪, ORCID: 0000-0003-2932-0342, E-mail: qxie@cse.cuhk.edu.hk。
*本文系国家自然科学基金项目“基于非对称群体兴趣相关性并融合情境与群体信任的Web服务推荐研究”(项目编号:61502401)和国家自然科学基金项目“基于切片技术的可信软件开发过程的B方法的研究”(项目编号:61379019)的研究成果之一。

高推荐的性能。

此外，另一部分研究工作使用聚类算法对相关用户进行聚类，再根据聚类结果进行 Web 服务的推荐[3, 8]。这些聚类算法与本文所用的相似度距离方法的区别是不仅群体划分的依据不同，而且相似性计算的顺序和方法不同。使用聚类算法的 Web 服务推荐使用用户的情境信息(如：地理位置)进行群体的划分，之后再计算单个用户与用户群体之间的相似性，此外情境信息的缺失会导致这些聚类算法的失效。本文方法仅使用用户调用服务的 QoS 矩阵而不依赖于特殊的情境数据。首先计算单个用户(服务)之间的相似性，然后依据单个目标用户(服务)的相似性距离为用户(服务)构建用户群体(服务群体)，再计算用户群体(服务群体)之间的相似性，从而更好地度量用户之间或服务之间的相似性。林耀进等[12]在传统推荐领域内提出一种基于用户群体影响的协同过滤推荐算法。本文与该算法不同之处在于：应用场景不同，不同于传统的对商品评分进行协同过滤推荐，本文是应用于服务计算领域对 Web 服务质量进行预测和推荐；同时使用用户和服务群体中心之间的相似性构建新的混合型推荐算法；在群体预测公式中以群体中心均值代替文献[12]中用户的均值；不仅考虑用户群体之间的关系而且将服务群体之间的关系也引入 Web 服务推荐，充分利用基于用户群体和基于服务群体的推荐优势，提出一种基于相似性群体的混合型 Web 服务质量推荐算法(Similarity Group based Hybrid Web Service QoS Recommendation Algorithm, GHQR)。该算法利用群体代替个体用户或服务，从而可以有效地解决 Web 服务推荐中数据稀疏性导致的相似服务或相似用户缺失的问题，在提高 Web 服务推荐准确性的同时获得较高的预测值覆盖率。

2 基于相似性群体的混合型 Web 服务推荐算法

为了解决 Web 服务推荐算法数据稀疏性问题的同时提高推荐算法的性能，GHQR 充分考虑用户群体和服务群体对 Web 服务推荐结果的影响，并依据相似性距离构建相应的用户群体和服务群体。GHQR 主要包括以下模块：用户群体和服务群体的构建、用户群体和服务群体相似性计算、相似群体邻居的选择、混合型 Web 服务推荐公式、算法及性能分析。

2.1 预备知识

基于协同过滤的 Web 服务推荐主要使用基于用户的协同过滤技术和基于项目的协同过滤技术[1-3,5,9]。其中，相似性的计算方法大多采用皮尔森相关系数法(Pearson Correlation Coefficient, PCC)[2]。基于用户的皮尔森相关系数(User-based Pearson Correlation Coefficient, UPCC)和基于项目的皮尔森相关系数(Item-based Pearson Correlation Coefficient, IPCC)分别计算用户之间的相似性和项目之间的相似性[2]。

2.2 用户群体和服务群体的构建

受到文献[12]的启发，用户的群体影响能更加精确地刻画用户之间的相似度。本文将文献[12]的用户群体方法应用于服务计算领域，并分别定义本文的用户群体和服务群体。

定义 1 给定用户集合 $U=\{u_1, u_2, \cdots, u_m\}$，为 U 中的任意用户 u_i，定义其α群体为：$\alpha(u_i)=\{\mu \mid \mu \in U, \Delta(\mu,u_i) \leqslant 1-\alpha\}-\{u_i\}$，其中 $\Delta(\mu,u_i)$ 为空间度量，$\alpha(u_i)$表示为服务用户 u_i 构建的用户群体α，并且相似度距离阈值α取值为 $0 \leqslant \alpha \leqslant 1$[12]。

定义 2 给定服务集合 $S=\{s_1, s_2, \cdots, s_n\}$，为 S 中的任意服务 s_i，定义其β群体为：$\beta(s_i)=\{\psi \mid \psi \in S, \Delta(\psi,s_i) \leqslant 1-\beta\}-\{s_i\}$，其中 $\Delta(\psi,s_i)$ 为空间度量，$\beta(s_i)$ 表示为服务用户 s_i构建的服务群体β，并且相似度距离阈值β取值为 $0 \leqslant \beta \leqslant 1$[12]。

根据定义 1 和定义 2，可以构建基于相似度距离阈值α的用户群体$\alpha(u_i)$和基于相似度距离阈值β的服务群体$\beta(s_i)$。当距离阈值α和距离阈值β变小时，对应的空间度量值将增大，用户群体和服务群体中包括的用户和服务数量会随之增加。用户之间和服务之间的差别将减少。由此可见，用户群体和服务群体的大小将会影响用户之间和服务之间的相似关系。

2.3 用户群体和服务群体相似性计算

不同于林耀进等[12]使用平均相似性方法计算用户群体的相似性，本文考虑到用户群体和服务群体的群体特征，定义群体中心代替群体，从而计算用户群体和服务群体之间的相似性。

定义 3 用户群体中心 $p_i=\{\overline{r_{p_i,s_1}}, \overline{r_{p_i,s_2}}, \cdots, \overline{r_{p_i,s_n}}\}$ 被定义为用户群体$\alpha(u_i)$中所有用户调用服务向量的平均 QoS 向量。

定义 4 服务群体中心 $q_i=\{\overline{r_{u_1,q_i}},\overline{r_{u_2,q_i}},\cdots,\overline{r_{u_m,q_i}}\}$ 被定义为在服务群体$\beta(s_i)$中用户调用所有 Web 服务的平均 QoS 向量。

根据定义 3 和定义 4, 设计用户群体和服务群体之间的相似性计算公式如下:

$$\mathrm{UGSim}(p_1,p_2)=\frac{\sum_{j\in I(p_1)\cap I(p_2)}(r_{p_1,j}-\overline{r_{p_1}})(r_{p_2,j}-\overline{r_{p_2}})}{\sqrt{\sum_{j\in I(p_1)\cap I(p_2)}(r_{p_1,j}-\overline{r_{p_1}})^2}\sqrt{\sum_{j\in I(p_1)\cap I(p_2)}(r_{p_2,j}-\overline{r_{p_2}})^2}} \quad (1)$$

$$\mathrm{SGSim}(q_1,q_2)=\frac{\sum_{u\in U(q_1)\cap U(q_2)}(r_{u,q_1}-\overline{r_{q_1}})(r_{u,q_2}-\overline{r_{q_2}})}{\sqrt{\sum_{u\in U(q_1)\cap U(q_2)}(r_{u,q_1}-\overline{r_{q_1}})}\sqrt{\sum_{u\in U(q_1)\cap U(q_2)}(r_{u,q_2}-\overline{r_{q_2}})}} \quad (2)$$

其中, $\mathrm{UGSim}(p_1, p_2)$为用户群体 p_1 和用户群体 p_2 的群体相似性值, $\mathrm{SGSim}(q_1, q_2)$为服务群体 q_1 和服务群体 q_2 的群体相似性值。j 属于用户群体 p_1 和用户群体 p_2 共同调用的服务集合, u 属于同时调用服务群体 q_1 和服务群体 q_2 的用户集合。

2.4 相似群体邻居的选择

传统使用 Top-K 算法对相似邻居的相似性进行排序, 从而选择相似性较高的 K 个相似邻居进行服务质量的预测[2]。为了提高预测的准确性, 笔者提出以群体邻居代替单个用户或服务邻居进行 Top-K 的选择, 其公式定义如下:

$$\mathrm{TUGSim}(p)=\{sp\mid sp\in \mathrm{TopK}(p),\mathrm{UGSim}(sp,p)>0,sp\neq p\} \quad (3)$$

$$\mathrm{TSGSim}(q)=\{sq\mid sq\in \mathrm{TopK}(q),\mathrm{SGSim}(sq,q)>0,sq\neq q\} \quad (4)$$

其中, TopK(p)表示用户群体 p 的最相似 K 个用户群体集合, TopK(q)表示服务群体 q 的最相似性 K 个服务群体集合。

2.5 混合型 Web 服务推荐公式

本文提出的 GHQR 将基于用户群体的推荐和基于服务群体的推荐相结合, 以解决 QoS 数据稀疏性问题并提高推荐的预测准确性。混合型 Web 服务推荐公式定义如下:

$$\begin{aligned}\mathrm{USGP}(r_{u,j})=&\gamma\times(\overline{r_p}+\frac{\sum_{sp\in \mathrm{TUGSim}(p)}\mathrm{UGSim}(p,sp)\times(r_{sp,j}-\overline{r_{sp}})}{\sum_{sp\in \mathrm{TUGSim}(p)}\mathrm{Sim}(p,sp)})+\\&(1-\gamma)(\overline{r_j}+\frac{\sum_{sq\in \mathrm{TSGSim}(q)}\mathrm{SGSim}(q,sq)\times(r_{u,sq}-\overline{r_{sq}})}{\sum_{sq\in \mathrm{TSGSim}(q)}\mathrm{SGSim}(q,sq)})\end{aligned} \quad (5)$$

其中, 参数γ和参数$(1-\gamma)$分别为基于用户群体预测方法和基于服务群体预测方法在混合型预测算法中所占的预测权重。p 为用户 u 所在的用户群体中心, sp 属于用户群体中心 p 的 Top-K 相似用户群体中心集合; q 为服务 j 所在的服务群体中心, sq 属于服务群体中心 q 的 Top-K 相似服务群体中心集合。$\overline{r_p}$ 表示 u 所在的用户群体中心 p 对调用的 Web 服务所获得的 QoS 值均值。

在混合型 Web 服务推荐公式中, 基于用户群体的预测方法使用用户群体中心均值代替传统算法中的用户均值, 而对于基于服务群体的预测方法仍然保留传统算法中的服务均值方法。这是由于从定义 4 中可以看出, 服务群体中心实质是用户的 QoS 平均值, 而用户平均值的预测准确性会低于服务平均值, 因此为了提高 Web 服务的推荐准确性, 保持原有算法的预测优势, 在混合型 Web 服务推荐算法中保留服务群体计算过程中的服务平均方法($\overline{r_j}$), 相似性的计算则仍改进为基于服务群体中心的算法。

2.6 GHQR 算法及性能分析

输入: QoS 数据的训练集 TrainingData, QoS 数据的测试集 TestData, 服务用户数 M, Web 服务数 N, 相似邻居数 Top-K, 相似度距离阈值α, 相似度距离阈值β

输出: $P(r_{u,j})$

Begin:

①SimU=ComputeSimU (TrainingData, M);
 SimS=ComputeSimS (TrainingData, N);

②GroupCenterMatrixU=
  ConstructGroupCenterMatrixU(SimU, α, M, N);
 GroupCenterMatrixS=
  ConstructGroupCenterMatrixS(SimS, β, M, N);

③GroupSimU= ComputeSimU(GroupCenterMatrixU, M);
 GroupSimS= ComputeSimS(GroupCenterMatrixS, N);

④**for** each missing $r_{u,j}$ in TestData **do**

⑤Sort the similarity values for user group of u-th row in GroupSimU with descend;

⑥Choose top K user group centers from the sorted user group similarities;

⑦Sort the similarity values for service group of j-th row in GroupSimS with descend;

⑧Choose top K service group centers from the sorted service group similarities;

⑨$P(r_{u,j})$ = PredictionQoSValue (TestData, GroupCenterMatrixU, GroupCenterMatrixS, K);

⑩**end for**

⑪**for** each target service user **do**

⑫Recommend a list of the functionally-equivalent candidate Web

services which have high predicted QoS values in $P(r_{u,j})$ for target user;

⑬**end for**

End

在算法中, ComputeSimU 和 ComputeSimS 函数分别使用皮尔森相关系数法计算用户和服务之间的相似性; 根据定义 1 和定义 2, 使用 ConstructGroupCenterMatrixU 和 ConstructGroupCenterMatrixU 函数构建用户和服务群体中心矩阵, 然后计算用户和服务群体中心之间的相似性; 为目标用户分别选择 Top-K 用户和服务群体中心的相似近邻; 运用公式(5)使用 PredictionQoSValue 函数计算缺省的 QoS 预测值; 最后根据预测的 QoS 值为目标用户推荐具有最优 QoS 预测值的 Web 服务列表以满足服务用户的需求。

在相似性计算时间复杂度上, GHQR 算法与传统的基于用户的协同推荐($O(M^2)$)或基于服务的协同推荐($O(N^2)$)算法相比, GHQR 增加了步骤③计算群体中心之间的相似性。由于基于用户和基于服务的推荐算法可以并行计算, 故 GHQR 的时间复杂度为 $\min(O(2M^2), O(2N^2))$。由此可见, 与传统协同过滤推荐算法相比 GHQR 并未增加算法的时间复杂度。

3　实　验

3.1　数据来源

为了验证 GHQR 算法的有效性和准确性, 采用拥有大规模真实 Web 服务质量调用记录的数据集 WS-DREAM①。该数据集包含分布于世界各地的 339 个服务用户对 5 825 个 Web 服务调用的 1 974 675 条 QoS 值记录[2]。本实验使用响应时间(Response Time, RT)进行对比实验, WS-DREAM 中部分响应时间(单位为秒)的数据如表 1 所示, 矩阵中的数据表示用户 u_i 调用服务 s_j 的响应时间值。

表 1　WS-DREAM 中部分响应时间数据

	s_1	s_2	s_3	s_4
u_1	5.982	0.222	5.776	0.329
u_2	2.130	0.254	0.740	0.244
u_3	0.854	0.358	1.271	0.211
u_4	0.693	0.219	0.960	0.304

①http://www.wsdream.net/.

3.2　算法推荐标准

为了评估算法的预测性能, 使用的评价指标包括标准平均绝对误差(Normalized Mean Absolute Error, NMAE)和覆盖率(Coverage)。其中 NMAE 的定义公式如下所示:

$$NMAE = \frac{\sum_{u,j}|r_{u,j} - \hat{r}_{u,j}|}{\sum_{u,j} r_{u,j}} \tag{6}$$

其中 $r_{u,j}$ 和 $\hat{r}_{u,j}$ 分别为用户 u 对服务 j 调用的真实响应时间值和预测响应时间值。从 NMAE 的定义可以看出, NMAE 值越小, 表明预测的响应时间值与真实的响应时间值之间的误差越小, 从而获得的预测准确性越高。

针对 Web 服务响应时间值的属性, 覆盖率定义为预测响应时间值大于 0 的预测服务数量(plusNum)与实际需要预测的全部服务数量(Num)的百分比, 覆盖率值越大表明预测算法能正确预测 Web 服务 QoS 值的百分比越高, 从而表明预测结果越好。具体计算如下:

$$Coverage = \frac{plusNum}{Num} \tag{7}$$

3.3　实验结果及分析

在本实验中, 笔者将 GHQR 算法与 4 种代表性推荐算法作对比分析: 传统的基于用户的协同过滤推荐算法(UPCC)、基于项目的协同过滤推荐算法(IPCC)、著名的基于协同过滤 QoS 感知的 Web 服务推荐算法(WSRec)[2]和基于用户群体影响的协同过滤推荐算法[12](本文简称 CFBUGI)。

为了构造真实 Web 服务数据稀疏性的场景, 从 WS-DREAM 数据集 5 825 个 Web 服务中任意挑选出不同数量的 Web 服务, 使用参数 χ 控制挑选的 Web 服务数量, 并设置 $\chi=1000$ 以构造数据稀疏性的场景。将包含 $M\times\chi$ 的数据集(命名为 RTset)分为两个部分, 一部分为预测算法的训练集; 另一部分为预测算法的测试集。参数 Percent 表示训练集所占 RTset 的百分比。此外, 使用变量 Density 设置训练集的稀疏性。变量 Top-K 为选择的相似邻居的数量。

表 2 展示了 GHQR 算法与 4 种推荐算法的 NMAE 和 Coverage 值对比结果。其中, 实验环境设置为

Percent=80%，Top-K=30，群体 α=0.8，群体 β=0.8，Density 分别为 10%、20%和 30%，相应的γ值设置为 γ=0.8，γ=0.65，γ=0.45。对推荐算法 WSRec 设置 Top-K=90, UPCC 和 IPCC 的权重分别为 80%和 20%。从表 2 可以看出随着 Density 从 10%递增到 30%, GHQR 均获得最低的 NMAE 值和最高的 Coverage 值。表 3 是 GHQR 算法与其余 4 种算法相比在 NMAE 和 Coverage 上获得的提高百分比。其中 NMAE 和 Coverage 获得的最优百分比分别为 69%和 188%。如表 2 和表 3 所示，表明 GHQR 与其余 4 种方法相比在不同稀疏性的设置下均获得最优的预测准确性，同时获得最高的预测值覆盖率，从而验证了 GHQR 算法的优越性。

表 2　标准平均绝对误差和覆盖率对比结果

Metric	Density	UPCC	IPCC	WSRec	CFBUGI	GHQR
NMAE	10%	1.06642	2.50279	1.06642	0.81416	**0.72538**
	20%	1.04139	2.26790	0.95167	0.77757	**0.73229**
	30%	1.04981	2.27672	0.97152	0.78899	**0.73896**
Coverage	10%	0.48334	0.27345	0.48334	0.79175	**0.98190**
	20%	0.43179	0.39325	0.25057	0.83998	**0.91787**
	30%	0.41421	0.41814	0.27807	0.80869	**0.82027**

表 3　GHQR 算法在标准平均绝对误差和覆盖率上的提高

Metric	Density	Impro. Vs UPCC	Impro. Vs IPCC	Impro. Vs WSRec	Impro. Vs CFBUGI
NMAE	10%	32%	71%	32%	10.9%
	20%	30%	68%	23%	5.8%
	30%	30%	68%	24%	6.3%
Average	–	**31%**	**69%**	**26%**	**7.7%**
Coverage	10%	103%	259%	103%	2.4%
	20%	113%	133%	266%	9.3%
	30%	98%	96%	195%	1.4%
Average	–	**105%**	**163%**	**188%**	**4%**

这是因为在数据稀疏性的条件下，使用单个用户(服务)的推荐算法可能会面临相似邻居缺失的问题或其相似邻居实质上并不太相似，而是刚好与目标用户(服务)共同调用过较少数量的服务(用户)[2]。然而相似邻居的寻找是协同过滤推荐算法中最重要的步骤之一。在相似邻居不具有高相似性的情况下进行推荐可能导致预测 QoS 值与真实 QoS 值之间具有较大的误差。因此推荐算法的预测准确性和覆盖率都需获得进一步的提高。如表 1 中以单个用户(服务)进行推荐的算法(UPCC，IPCC，WSRec)获得较低的预测准确性和覆盖率。以群体代替单个用户(服务)的推荐算法，是使用用户(服务)群体中心代替单个用户(服务)，计算用户(服务)群体与用户(服务)群体之间的群体相似性代替单个用户(服务)与单个用户(服务)之间的相似性，从而根据群体相似性寻找到更准确的群体相似邻居进行推荐，以获得更准确的预测 QoS 值，从而获得较高的预测准确性和覆盖率。以群体进行推荐的算法(CFBUGI、GHQR)获得较优的预测性能从而验证了上述的推理。此外，在本文所提算法中对典型的混合型算法进行改进，并充分利用原有基于用户和基于服务推荐算法的预测优势，从而与 GHQR 群体算法相比具有较优的预测性能。

(1) 预测权重 γ 对预测结果的影响

为了研究预测权重 γ 对预测结果的影响，将实验环境设置为 Percent=80%，K=30，群体 α=0.8，群体 β=0.8, Density 分别为 10%、20%和 30%。由于当预测权重 γ 为 0 时混合算法的预测结果仅为基于用户群体的预测结果，而当预测权重 γ 为 1 时，混合算法的预测结果仅为基于服务群体的预测结果。所以将预测权重 γ 以步长 0.05 从初始值 0.05 递增到 0.95 以研究 γ 对 GHQR 预测结果的影响。

图 1 和图 2 分别为 GHQR 与基于用户群体影响的协同过滤算法 CFBUGI 在 NMAE 和 Coverage 指标上的预测性能对比。从图 1 中可以看出，不同稀疏性情况下当γ值增加的时候，首先 GHQR 的 NMAE 值持续降低，到达最低值时再增加。这表明在不同稀疏性情况下 GHQR 的最佳值有所不同，并且随着γ值增加基于用户群体算法在 GHQR 中的权重值逐步下降。当稀疏性为 10%、20%、30%时，获得最佳预测值的γ分别为 0.8(0.72538)、0.65(0.73229)、0.45(0.73896)。与 CFBUGI 算法相比，不同稀疏性情况下的 GHQR 最佳 NMAE 值均小于 0.74，表明 GHQR 具有较优的预测准确性。同理从图 2 中可以看出，随着预测权重γ值的不断增加，GHQR 的 Coverage 值也持续增加并最后到达平稳状态。当稀疏性为 10%、20%、30%时，与 CFBUGI 的 Coverage 值 0.80869、0.83998、0.79175 相比，GHQR 获得的最佳 Coverage 值分别为 0.99286、0.98001、0.86577。从而表明 GHQR 算法具有较优的预测性能。

图 1 预测权重γ对 NMAE 的影响

图 2 预测权重γ对 Coverage 的影响

(2) 群体相似度阈值α(β)对预测结果的影响

根据定义 1 和定义 2，群体相似度距离阈值α和β分别控制群体用户和群体服务的空间大小。为了研究这两个参数对预测结果的影响，如图 3 所示，将实验环境设置为 Percent=80%，K=30，Density=10，γ =0.6。群体相似度阈值α和群体β取相同的值，标记为α(β)。α和β值以 0.1 为步长从 0 递增到 1。与基于用户群体影响的协同过滤算法 CFBUGI 相比，GHQR 在不同α和β取值下，均获得较低的 NMAE 值和较高的 Coverage 值，从而表明 GHQR 不仅能提高预测的准确性，而且在预测值覆盖率上获得显著提高。

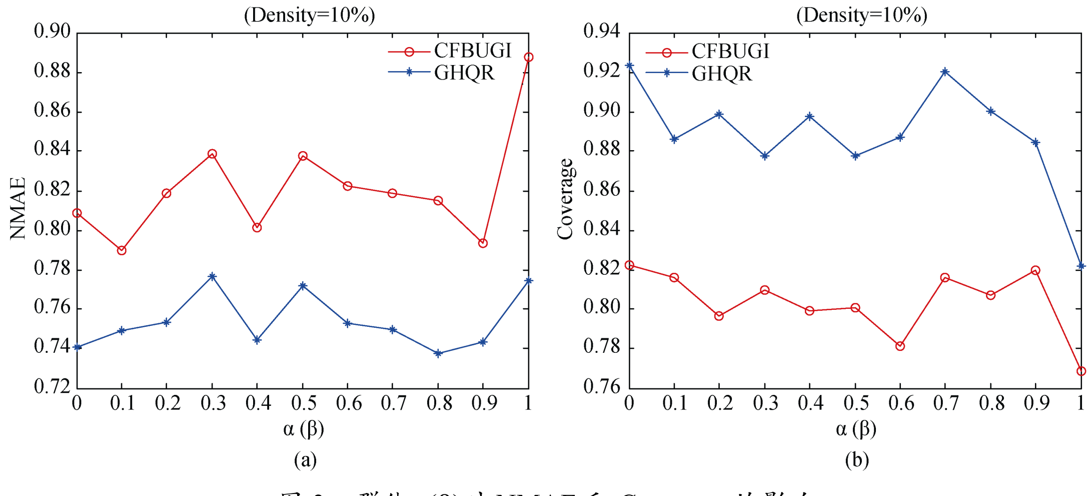


图 3 群体α(β)对 NMAE 和 Coverage 的影响

4 结 语

本文利用群体特性提出一种基于相似性群体的混合型 Web 服务推荐算法，该算法充分利用基于用户群体和基于服务群体的优点，分别为目标用户和服务根据相似性距离构造个性化的相似性用户群体和服务群体，定义用户群体和服务群体中心，并设计新的融合用户群体中心和服务群体中心之间相似性的混合型 Web 服务推荐算法。实验结果表明本文所提出的预测算法不仅能提高 Web 服务预测的准确性而且能获得较高的覆盖率。在后续的工作中将对更多的 QoS 属性进行研究以拓展混合型算法的应用场景。对 Web 服务 QoS 值预测准确性有影响的其他因素(如: QoS 的时序性, 可信度等)也将是后续工作的研究重点。

参考文献:

[1] 邵凌霜, 周立, 赵俊峰, 等. 一种 Web Service 的服务质量预测方法[J]. 软件学报, 2009, 20(8): 2062-2073. (Shao Lingshuang, Zhou Li, Zhao Junfeng, et al. Web Service QoS Prediction Approach [J]. Journal of Software, 2009, 20(8): 2062-2073.)

[2] Zheng Z, Ma H, Lyu M R, et al. QoS-aware Web Service Recommendation by Collaborative Filtering [J]. IEEE Transactions on Service Computing, 2011, 4(2): 140-152.

[3] Chen X, Zheng Z, Liu X, et al. Personalized QoS-aware Web Service Recommendation and Visualization [J]. IEEE Transactions on Services Computing, 2013, 6(1): 35-47.

[4] He P, Zhu J, Zheng Z, et al. Location-based Hierarchical Matrix Factorization for Web Service Recommendation [C]. In: Proceedings of 2014 IEEE International Conference on Web Services (ICWS2014). Washington, DC: IEEE Computer Society, 2014: 297-304.

[5] 张璇, 刘聪, 王黎霞, 等. 基于协同过滤的可信 Web 服务推荐方法[J]. 计算机应用, 2014, 34(1): 213-217. (Zhang Xuan, Liu Cong, Wang Lixia, et al. Trustworthy Web Service Recommendation Based on Collaborative Filtering [J]. Journal of Computer Applications, 2014, 34(1): 213-217.)

[6] 王海艳, 杨文彬, 王随昌, 等. 基于可信联盟的服务推荐方法[J]. 计算机学报, 2014, 37(2): 301-311. (Wang Haiyan, Yang Wenbin, Wang Suichang, et al. A Service Recommendation Method Based on Trustworthy Community [J]. Chinese Journal of Computers, 2014, 37(2): 301-311.)

[7] 姜波, 张晓筱, 潘伟丰. 基于二部图的服务推荐算法研究[J]. 华中科技大学学报: 自然科学版, 2013, 41(S2): 93-99. (Jiang Bo, Zhang Xiaoxiao, Pan Weifeng. Bipartite Graph-based Service Recommendation Method Study [J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2013, 41(S2): 93-99.)

[8] 林祥云, 刘小青, 唐明董, 等. Web 服务 QoS 与用户位置的相关性实证研究[J]. 计算机工程与科学, 2013, 35(9): 83-88. (Lin Xiangyun, Liu Xiaoqing, Tang Mingdong, et al. An Empirical Study of Correlation Between Web Service QoS and User Location [J]. Computer Engineering and Science, 2013, 35(9): 83-88.)

[9] Xie Q, Zheng Z, Liu L, et al. Correlation-based Top-k Recommendation for Web Services [C]. In: Proceedings of the 13th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC-2015), Liverpool, UK. 2015: 1903-1909.

[10] 高虎明, 赵凤跃. 一种融合协同过滤和内容过滤的混合推荐方法[J]. 现代图书情报技术, 2015(6): 20-26. (Gao Huming, Zhao Fengyue. A Hybrid Recommendation Method Combining Collaborative Filtering and Content Filtering [J]. New Technology of Library and Information Service, 2015(6): 20-26.)

[11] 祝婷, 秦春秀, 李祖海. 基于用户分类的协同过滤个性化推荐方法研究[J]. 现代图书情报技术, 2015(6): 13-19. (Zhu Ting, Qin Chunxiu, Li Zuhai. Research on Collaborative Filtering Personalized Recommendation Method Based on User Classification [J]. New Technology of Library and Information Service, 2015(6): 13-19.)

[12] 林耀进, 胡学钢, 李慧宗. 基于用户群体影响的协同过滤推荐算法[J]. 情报学报, 2013, 32(3): 299-305. (Lin Yaojin, Hu Xuegang, Li Huizong. Collaborative Filtering Recommendation Algorithm Based on User Group Influence [J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(3): 299-305.)

[13] 盈艳, 曹妍, 牟向伟. 基于项目评分预测的混合式协同过滤推荐[J]. 现代图书情报技术, 2015(6): 27-32. (Ying Yan, Cao Yan, Mu Xiangwei. A Hybrid Collaborative Filtering Recommender Based on Item Rating Prediction [J]. New Technology of Library and Information Service, 2015(6): 27-32.)

作者贡献声明:

谢琪, 崔梦天: 提出研究思路, 设计研究方案, 论文起草和最终版本修订;

谢琪: 数据的获取、提供与分析, 进行实验。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: zibin.gil@gmail.com。

[1] 郑子彬. 339-5825-QoSDataset2.zip. 339 个用户调用 5 825 个服务的 QoS 调用记录.

收稿日期: 2016-03-07
收修改稿日期: 2016-04-14

Group Similarity Based Hybrid Web Service Recommendation Algorithm

Xie Qi[1,2]　Cui Mengtian[1]
[1](School of Computer Science and Technology, Southwest University for Nationalities, Chengdu 610225, China)
[2](Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong 999077, China)

Abstract: [**Objective**] This paper tries to solve the issues of lacking similar services or users in Web service computing due to the data sparsity of Quality of Service (QoS) recommendation. **[Methods]** First, we created personalized similar user and service groups according to similarity distance of the target users and services. Second, we used the group center similarities of the user and service groups to design a new hybrid recommendation algorithm(GHQR), which was tested with real-world data of 1.97 million QoS records. **[Results]** Compared with two traditional recommendation algorithms, the GHQR reduced the Normalized Mean Absolute Error (NMAE) by 31% and 69%. It also increased the Coverage by 105% and 163%, respectively. **[Limitations]** Our study only examined the response time of QoS, and more research was needed to investigate other QoS properties. **[Conclusions]** Comprared with WSRec and CFBUGI, the GHQR can reduce the NMAE by 26% and 7.7%. It also increased the Coverage by 188% and 4%, respectively. GHQR not only enhances the prediction accuracy but also increases the coverage significantly.

Keywords: Personalized recommendation　Service computing　Web services　Quality of Service